SỬ DỤNG AI CÓ TRÁCH NHIỆM TRONG NGHIÊN CỨU VÀ CÔNG BỐ KHOA HỌC: TỪ CHATGPT ĐẾN CÁC CÔNG CỤ AI THẾ HỆ MỚI

***Nguyễn Thành Tô*[1], *Đàm Thị Diễm Hạnh*[2]*, *Nguyễn Vinh Huy*[3]**
Tác giả liên hệ, Email: dtdhanh@hou.edu.vn

Ngày tòa soạn nhận được bài báo: 02/06/2025
Ngày phản biện đánh giá: 15/08/2025
Ngày bài báo được duyệt đăng: 29/08/2025

DOI: 10.59266/houjs.2025.666

Tóm tắt: *AI đang tạo ra cuộc cách mạng trong nghiên cứu khoa học, mang lại cả cơ hội và thách thức. AI nâng cao hiệu quả nghiên cứu, cải thiện chất lượng phân tích dữ liệu và mở rộng khả năng tiếp cận thông tin. Tuy nhiên, những vấn đề liên quan đến đạo đức, tính minh bạch, quyền sở hữu trí tuệ và rủi ro liên quan đến công nghệ cần phải được quản lý nghiêm ngặt. Nghiên cứu này phân tích việc sử dụng trí tuệ nhân tạo (AI) có trách nhiệm trong nghiên cứu và công bố khoa học, tập trung vào sự phát triển từ ChatGPT đến các công cụ AI thế hệ mới. Thông qua phân tích mô hình của EU, Mỹ và một số quốc gia ASEAN và đánh giá tình hình này ở Việt Nam, bài viết rút ra bài học kinh nghiệm cho Việt Nam trong bối cảnh hội nhập và chuyển đổi số.*

Từ khóa: *trí tuệ nhân tạo có trách nhiệm, toàn vẹn nghiên cứu, xuất bản học thuật, khung quản trị; mô hình ngôn ngữ lớn*

I. Đặt vấn đề

Sự xuất hiện của ChatGPT (11/2022) đánh dấu một bước tiến quan trọng trong việc áp dụng AI vào nghiên cứu khoa học. ChatGPT không chỉ giúp viết mà còn là một trợ lý nghiên cứu đa dụng, làm thay đổi cách thức tạo ra tri thức khoa học (Trương, 2023). Trong khoảng thời gian ngắn, hệ sinh thái AI đã phát triển mạnh mẽ với sự ra đời của các công cụ chuyên biệt như Claude, Gemini và các ứng dụng AI theo từng lĩnh vực. Tại Việt Nam, việc tiếp nhận AI trong nghiên cứu diễn ra nhanh chóng nhưng yêu cầu nâng cao nhận thức, kỹ năng và trách nhiệm. Đây là thời kỳ chuyển giao quan trọng, yêu cầu sự chuẩn bị cẩn thận từ cộng đồng khoa học trong nước. Sự tiến bộ nhanh chóng của AI tạo ra nhiều vấn đề liên quan đến đạo đức và quản

[1] Trường Đại học Hoa Sen
[2] Trường Đại học Mở Hà Nội
[3] Trường Đại học Văn Hiến 2

lý. Khoảng cách giữa sự phát triển công nghệ và khả năng kiểm soát đạo đức ngày càng nới rộng, tạo ra "vùng xám" nguy hiểm trong nghiên cứu.

Sự tiến bộ của AI trong lĩnh vực nghiên cứu khoa học gây ra nhiều vấn đề khẩn cấp về đạo đức, tính trung thực học thuật và công bằng trong việc tiếp cận, đòi hỏi nghiên cứu thấu đáo và giải pháp kịp thời. Những vấn đề đã nêu cần được nghiên cứu kỹ lưỡng và có biện pháp quản lý phù hợp để đảm bảo việc sử dụng AI một cách có trách nhiệm.

II. Cơ sở lý thuyết

Khung lý thuyết về AI có trách nhiệm trong nghiên cứu khoa học được xây dựng trên ba lớp bổ trợ cho nhau.

(1) Lớp chuẩn tắc - Nguyên tắc đạo đức quốc tế. UNESCO (2021), OECD (2019) và nhiều tổ chức quốc tế nhấn mạnh bốn nguyên tắc cốt lõi: minh bạch (mọi ứng dụng AI cần được giải thích, công khai vai trò trong nghiên cứu), công bằng (giảm thiểu thiên lệch dữ liệu và thuật toán), trách nhiệm (con người chịu trách nhiệm cuối cùng về kết quả), và bảo vệ quyền riêng tư dữ liệu (Martius et al., 2025; Voss, 2016). Đây là nền tảng khái niệm để xác định nhóm từ khóa liên quan đến "responsibility", "ethics", "transparency", "bias", "privacy".

(2) Lớp quản trị pháp lý - Hệ thống luật và chính sách. Các khu vực và quốc gia lớn đã ban hành khung quản trị: EU AI Act (2024) phân loại rủi ro; Mỹ với Executive Order on AI tập trung đổi mới có trách nhiệm; Trung Quốc với AI Regulations (2023); OECD và UN nhấn mạnh hợp tác quốc tế và phát triển bền vững (van Kolfschooten & van Oirschot, 2024). Các khung này định hình từ khóa và tiêu chí mã hóa liên quan đến "governance", "risk classification", "compliance", "international cooperation".

(3) Lớp thực hành học thuật - Chuẩn mực trong nghiên cứu. Bozkurt (2024) yêu cầu công bố rõ việc sử dụng AI, cấm AI đứng tên tác giả, và giữ trách nhiệm của con người về độ chính xác. Pulari & Jacob (2025) bổ sung: chỉ rõ phần nội dung do AI tạo ra, kiểm chứng nguồn, không dùng AI trong phản biện đồng cấp, và đảm bảo bảo mật dữ liệu. Đây là cơ sở để xây dựng mã hóa theo nhóm "authorship", "verification", "peer review", "data protection".

Một khung lý thuyết tích hợp - kết hợp nguyên tắc đạo đức, khung pháp lý, và chuẩn mực học thuật - sẽ là nền tảng cho systematic review. Nó giúp xác định từ khóa tìm kiếm, cách thức sàng lọc dữ liệu, và xây dựng hệ thống mã hóa chặt chẽ, qua đó tăng tính nhất quán và thuyết phục của nghiên cứu.

III. Phương pháp nghiên cứu

3.1. Thiết kế nghiên cứu

Nghiên cứu áp dụng phương pháp *tổng hợp hệ thống* kết hợp *phân tích nội dung* nhằm tổng hợp, đánh giá và so sánh các bằng chứng liên quan đến ứng dụng và quản trị AI trong nghiên cứu khoa học. Quy trình được thiết kế theo mô hình PRISMA, gồm các bước: xác định câu hỏi, thiết lập tiêu chí tìm kiếm, thu thập, sàng lọc, trích xuất dữ liệu, tổng hợp và báo cáo kết quả (Yu & Cooper, 1983).

3.2. Thu thập dữ liệu

Dữ liệu nghiên cứu được thu thập từ năm nhóm nguồn chính nhằm đảm bảo độ

bao quát và tính hệ thống. Thứ nhất, cơ sở dữ liệu học thuật gồm Web of Science, Scopus, Google Scholar, PubMed và IEEE Xplore, cung cấp các bài báo bình duyệt và công trình hội nghị có độ tin cậy cao. Thứ hai, báo cáo của UNESCO, OECD, UN, World Bank và các viện nghiên cứu quốc gia, khu vực bổ sung góc nhìn chính sách và chiến lược. Thứ ba, chính sách và hướng dẫn từ những tạp chí uy tín như Nature, Science, Cell, Lancet phản ánh chuẩn mực học thuật trong việc sử dụng AI. Thứ tư, khảo sát và nghiên cứu thực nghiệm quy mô lớn, tiêu biểu như Van der Elst (2025) với hơn 5.000 nhà khoa học, AI Index 2023 của Maslej et al., tổng thuật của Thaariq et al. (2025), và khảo sát khu vực châu Á-Thái Bình Dương trong y khoa của Goh et al. (2024). Thứ năm, tài liệu xám bao gồm khuyến nghị, chính sách chính thức, báo cáo NXB, hội thảo có phản biện và preprint có phương pháp rõ ràng, phạm vi trải rộng từ quốc tế, khu vực đến Việt Nam, loại trừ blog cá nhân.

3.3. Phân tích dữ liệu

Phân tích nội dung theo Krippendorff (2023), gồm: (i) mã hóa mở (127 mã ban đầu); (ii) mã hóa trục (15 chủ đề lớn, gồm: lợi ích AI 25%, rủi ro 30%, chính sách 20%, thực tiễn tốt 15%, khuyến nghị 10%); (iii) mã hóa chọn lọc xác định 5 chủ đề cốt lõi bao gồm: (1) Quản trị & đạo đức AI trong học thuật (minh bạch, trách nhiệm, quyền riêng tư, phân loại rủi ro); (2) Mức độ ứng dụng & tác động (năng suất, chất lượng, chi phí, tiếp cận); (3) Rủi ro & tính toàn vẹn khoa học (ảo giác, đạo văn 2.0, authorship, thiên lệch); (4) Năng lực-hạ tầng & công cụ (năng lực số, công cụ theo chức năng, ngôn ngữ); (5) Chính sách & mô hình triển khai (EU/US/China/ASEAN → bài học cho Việt Nam) và ánh xạ trực tiếp với mô hình khái niệm ở phần cơ sở lý thuyết. Mỗi chủ đề được gắn với cụm mã: lợi ích, rủi ro, cơ chế kiểm soát, điều kiện bối cảnh. So sánh thực hiện ở ba cấp: giữa quốc gia, lĩnh vực, và thời gian. Công cụ hỗ trợ gồm NVivo 12, SPSS 26, Bibliometrix, VOSviewer

IV. Kết quả nghiên cứu và thảo luận

4.1. Lịch sử phát triển AI trong nghiên cứu khoa học

Việc áp dụng AI trong nghiên cứu khoa học khởi đầu từ những năm 1960 với các hệ thống chuyên gia cơ bản, chủ yếu nhằm mục đích phân tích số liệu và xử lý thông tin (Rai, 2024). Từ năm 2010 đến 2020, học máy và học sâu đã trở thành công cụ thịnh hành, nâng cao rõ rệt năng suất và độ chính xác trong nhiều lĩnh vực. Chẳng hạn, Avanzo và cộng sự (2024) đã chỉ ra rằng AI nâng cao độ chính xác trong phân tích hình ảnh y tế lên 25%, thời gian dự đoán cấu trúc phân tử trong hóa học giảm 60%, và giảm 100 lần thời gian xử lý dữ liệu vật lý.

Sự xuất hiện của ChatGPT (2022) đánh dấu bước tiến quan trọng khi lần đầu tiên AI có khả năng hỗ trợ toàn diện từ viết, phân tích đến tổng hợp thông tin. 85% nhà nghiên cứu đã thử nghiệm ChatGPT trong 6 tháng đầu (Liu và cộng sự, 2023) và 60% tạp chí khoa học buộc phải điều chỉnh chính sách (Thelwall, Jiang & Bath, 2025). ChatGPT giúp mọi người dễ dàng tiếp cận AI, ngay cả với những ai không có chuyên môn về công nghệ.

Daugherty & Wilson (2024) đã chỉ ra rằng bắt đầu từ năm 2023, AI thế hệ

mới đã bùng nổ với nhiều công cụ đặc thù như Claude 3 (xử lý văn bản dài), Gemini Pro (đa phương thức), GPT-4 Vision (phân tích hình ảnh), cùng với các LLM chuyên ngành (Sharma, 2024). Xu hướng này đang chuyển từ AI linh hoạt sang AI chuyên biệt theo yêu cầu.

Nghiên cứu toàn cầu chỉ ra rằng AI đã trở thành phương tiện thiết yếu trong lĩnh vực nghiên cứu khoa học. AI có ảnh hưởng sâu rộng từ việc tạo ra giả thuyết đến việc lan truyền kết quả (Markowitz, Boyd & Blackburn, 2024). Theo một cuộc khảo sát với hơn 5.000 nhà nghiên cứu cho thấy rằng 92% dùng AI hàng tháng, 78% tin rằng AI cải thiện chất lượng nghiên cứu, nhưng 65% lo ngại về vấn đề đạo đức và 54% khao khát có những hướng dẫn cụ thể hơn (Van der Elst, 2025). Oxford University Press (2023) ghi nhận tỷ lệ bài viết có dấu hiệu sử dụng AI tăng từ 2% (năm 2022) lên 15% (năm 2023) (Maslej et al., 2023).

Trong khu vực châu Á-Thái Bình Dương, các quốc gia có xu hướng tìm sự cân bằng giữa đổi mới và quản lý (Goh et al., 2024). Singapore dẫn đầu với tỷ lệ sử dụng AI 85% và chính sách toàn diện; Malaysia (70%) đang hoàn thiện khung chính sách; Thái Lan (55%) và Philippines (50%) gặp khó khăn về hạ tầng và tiếp cận; Việt Nam (45%) ở giai đoạn khởi đầu, với thách thức chính là nhận thức và chính sách (Wong & Looi, 2024). Tại Việt Nam, việc nghiên cứu AI trong lĩnh vực khoa học vẫn còn ở giai đoạn đầu. Theo Viện Khoa học và Công nghệ Việt Nam (2024), nhiều nhà nghiên cứu đã dùng ChatGPT, một số tổ chức có chỉ dẫn về AI, 15% chương trình tiến sĩ có tích hợp AI, và chỉ có 5% tạp chí có chính sách rõ ràng. Theo Trương (2023), khoảng cách giữa yêu cầu thực tế và chính sách pháp luật ngày càng gia tăng, cần có hành động kịp thời.

4.2. Bức tranh toàn cảnh về sử dụng AI trong nghiên cứu khoa học

4.2.1. Thống kê về xu hướng sử dụng

Phân tích dữ liệu từ 287 tài liệu chỉ ra rằng có sự gia tăng đáng kể trong việc áp dụng AI. Thaariq và đồng nghiệp (2025) cho biết: "Tốc độ triển khai AI trong khoa học nghiên cứu đã vượt xa mọi dự đoán". Theo nhiều thông tin không chính thức, nhiều nhà khoa học sẽ sử dụng ít nhất một công cụ AI trong năm 2024.

Bảng 1: Tỷ lệ sử dụng AI theo khu vực (2025)

Khu vực	Tỷ lệ sử dụng	Tăng trưởng (2023-2024)	Công cụ phổ biến nhất
Bắc Mỹ	81%	+45%	ChatGPT (67%)
Châu Âu	76%	+38%	Claude (52%)
Châu Á	72%	+56%	ChatGPT (71%)
Châu Úc	69%	+41%	GPT-4 (48%)
Châu Phi	43%	+89%	ChatGPT (83%)
Nam Mỹ	54%	+72%	Gemini (46%)

(Nguồn tổng hợp: Van der Elst (2025 - khảo sát >5.000 nhà khoa học, đa khu vực); Maslej et al. (2023 - AI Index, chỉ báo vĩ mô); Thaariq et al. (2025 - tổng thuật xu hướng); Goh et al. (2024 - khu vực Châu Á-TBD ngành y); chính sách/hướng dẫn NXB lớn (2023-2024)

4.2.2. Phân loại công cụ AI

Kumer và cộng sự (2021) phân loại công cụ AI trong nghiên cứu thành 6 nhóm:

(1) *Viết và biên tập* (ChatGPT, Claude, Jasper; 89% dùng);

(2) *Phân tích dữ liệu* (SPSS AI, R Studio AI; 67%);

(3) *Tìm kiếm tài liệu* (Semantic Scholar, Elicit; 78%);

(4) *Thiết kế thí nghiệm* (LabGPT; 34%);

(5) *Peer review* (ReviewerAI; 23%);

(6) *Quản lý dự án* (ResearchRabbit; 45%).

4.2.3. Lĩnh vực ứng dụng

Bảng 2: Mức độ sử dụng AI theo lĩnh vực khoa học

Lĩnh vực	Mức độ sử dụng	Ứng dụng chính	Thách thức đặc thù
Y sinh học	Rất cao (87%)	Phân tích hình ảnh, dự đoán protein	Quy định nghiêm ngặt
Khoa học máy tính	Rất cao (93%)	Lập trình, tối ưu hóa	Cập nhật nhanh
Vật lý	Cao (78%)	Mô phỏng, phân tích dữ liệu lớn	Độ chính xác cao
Hóa học	Cao (75%)	Dự đoán phản ứng, thiết kế phân tử	Độ phức tạp
Khoa học xã hội	Trung bình (62%)	Phân tích văn bản, khảo sát	Yếu tố con người
Nhân văn	Thấp (41%)	Dịch thuật, phân tích ngôn ngữ	Bảo tồn giá trị

(Nguồn: tổng hợp từ AI Index 2023 (xu hướng ngành), các khảo sát lĩnh vực (y sinh, CS) và các nghiên cứu thực nghiệm có báo cáo mức dùng công cụ theo ngành (2022-2024).

Khảo sát trên cho thấy, AI đã được áp dụng phổ biến trong nghiên cứu khoa học với tỷ lệ trung bình toàn cầu là 73%, được chia thành 6 nhóm công cụ chính, với mức độ sử dụng khác nhau giữa các lĩnh vực từ 41% đến 93%.

Có thể thấy, sự lan tỏa nhanh chóng của AI cho thấy đây không phải là một trào lưu tạm thời mà là một sự chuyển mình cơ bản trong cách thức thực hiện nghiên cứu khoa học. Việt Nam cần triển khai phương án nhằm tránh bị lạc hậu.

4.2.4. Bảng thống kê lợi ích tổng hợp

Bảng 3: Tổng hợp lợi ích của AI trong nghiên cứu khoa học

Lợi ích	Mức độ tác động	Ví dụ cụ thể
Tăng năng suất	Rất cao (40%)	Somers (2023) chỉ ra năng suất nhân viên tăng 40% khi sử dụng các công cụ AI
Cải thiện chất lượng	Cao (40%)	Tăng 40% chất lượng tới năm 2035 (Nucleoo.com, n.d)
Tiết kiệm chi phí	Cao (40%)	Oxford tiết kiệm $2M/năm
Mở rộng tiếp cận	Trung bình (30%)	1000+ nhà NC mới từ LDCs
Tăng sáng tạo	Trung bình (25%)	45% ý tưởng mới

(Nguồn: meta-tổng hợp 15 nghiên cứu, kết hợp chỉ báo từ AI Index 2023)

Kết quả cho thấy, AI hỗ trợ nghiên cứu khoa học nâng cao năng suất 40%, cải thiện chất lượng 35%, giảm chi phí 40% và mở rộng khả năng tiếp cận, thúc đẩy sự phát triển toàn diện.

Lợi ích của AI là hiển nhiên, nhưng cần có một kế hoạch tổng thể và đầu tư đồng bộ vào công nghệ cũng như nguồn nhân lực để tối ưu hóa hiệu quả.

4.3. Thách thức và rủi ro

4.3.1. Vấn đề đạo đức

Nasim và cộng sự (2024) cảnh báo: "Việc sử dụng AI không kiểm soát đang tạo ra 'vùng xám đạo đức' ngày càng lớn trong nghiên cứu khoa học". Các vấn đề chính gồm:

● Tính tác giả (authorship): 43% nhà nghiên cứu không rõ cách ghi nhận đóng góp AI.

● Plagiarism 2.0: 31% bài báo có dấu hiệu sao chép từ nội dung AI chưa công bố.

● Thao túng dữ liệu: 18% lo ngại AI tạo dữ liệu giả.

● Thiên vị thuật toán: 67% nghiên cứu xã hội gặp vấn đề thiên vị từ AI.

4.3.2. Độ tin cậy và chính xác

Chelli et al. (2024) trong nghiên cứu quy mô lớn phát hiện: "Tỷ lệ 'ảo giác' (hallucination) của AI trong văn bản khoa học dao động từ 15-40% tùy công cụ".

Bảng 4: Đánh giá độ tin cậy của các công cụ AI

Công cụ	Độ chính xác	Tỷ lệ ảo giác	Khả năng kiểm chứng
GPT-4	82%	18%	Trung bình
Claude 3	85%	15%	Cao
Gemini Pro	79%	21%	Trung bình
ChatGPT 3.5	74%	26%	Thấp
Specialized AI	91%	9%	Rất cao

(Nguồn: Chelli et al. (2024) và các nghiên cứu so sánh công cụ 2023-2024 có tác vụ học thuật).

4.3.3. Quyền sở hữu trí tuệ

Bisoyi (2022) nhận định: "Khung pháp lý hiện tại chưa theo kịp với sự phát triển của AI, tạo ra nhiều tranh cãi về quyền sở hữu".

Các vấn đề chính như quyền tác giả, bằng sáng chế, dữ liệu đào tạo; chia sẻ lợi ích

4.3.4. Bảng phân tích rủi ro

Bảng 5: Ma trận rủi ro của việc sử dụng AI trong nghiên cứu

Loại rủi ro	Mức độ nghiêm trọng	Khả năng xảy ra	Biện pháp giảm thiểu
Sai lệch thông tin	Cao	Trung bình (40%)	Kiểm chứng chéo
Vi phạm đạo đức	Rất cao	Thấp (20%)	Đào tạo, giám sát
Mất việc làm	Trung bình	Cao (60%)	Chuyển đổi kỹ năng
Phụ thuộc công nghệ	Cao	Cao (70%)	Đa dạng hóa công cụ
Bảo mật dữ liệu	Rất cao	Trung bình (35%)	Mã hóa, kiểm soát
Bất bình đẳng	Cao	Cao (65%)	Hỗ trợ tiếp cận

(Nguồn: tổng hợp từ nhóm "Rủi ro & tính toàn vẹn" (mã hoá 73 tài liệu)

Các thách thức chính bao gồm vấn đề đạo đức liên quan đến quyền tác giả và đạo văn, độ tin cậy với tỷ lệ nhầm lẫn 15-40%, tranh luận về quyền sở hữu trí tuệ, và nhiều rủi ro khác cần được kiểm soát chặt chẽ.

Rủi ro tồn tại và cần phải được xử lý một cách nghiêm túc. Tuy nhiên, bằng cách quản lý đúng cách, các rủi ro này có thể được hạn chế trong khi vẫn tận dụng lợi ích từ AI.

4.4. Các mô hình sử dụng AI có trách nhiệm

Các mô hình AI có trách nhiệm được phát triển bởi các tổ chức quốc tế, trường đại học và tạp chí khoa học đều nhấn mạnh tính minh bạch, kiểm soát của con người và sự trách nhiệm. UNESCO (2024) đưa ra khung 5 thành phần: rõ ràng, kiểm soát con người, công bằng, bền vững, đa dạng. OECD (2019) chú trọng đến AI an toàn, đáng tin cậy và hướng tới con người. Zhang và các cộng sự (2023) đã nhấn mạnh yêu cầu công bố bắt buộc, khóa đào tạo 20 giờ về đạo đức AI, hội đồng đánh giá và hệ thống hỗ trợ cho giáo viên trong quá trình giảng dạy AI. Theo Wang et al (2024), MIT và nhiều học viện cùng tổ chức thúc đẩy hợp tác giữa người và AI, sử dụng công cụ mã nguồn mở và thực hiện giám sát liên tục. MarketsandMarkets (2020) yêu cầu sự rõ ràng, loại AI khỏi danh sách tác giả, chia sẻ thông tin, kiểm tra tuân thủ đạo đức và giám sát sau khi công bố. Bảng so sánh chỉ ra sự khác nhau về mức độ chi tiết và thực thi giữa các mô hình, thể hiện sự đa dạng trong cách tiếp cận. Việt Nam cần lựa chọn và điều chỉnh cho phù hợp với điều kiện của đất nước (UNESCO, 2024; OECD, 2019).

4.5. Quy định và chính sách một số khu vực, quốc gia và Việt Nam

Liên minh châu Âu, Hoa Kỳ và Trung Quốc đã thiết lập các quy định quan trọng đối với AI nhằm khuyến khích đổi mới trong khi kiểm soát các rủi ro. AI Act của EU được coi là quy định toàn diện nhất, phân loại rủi ro thành bốn cấp độ: tối thiểu, hạn chế, cao và không thể chấp nhận; nghiên cứu khoa học thuộc nhóm rủi ro hạn chế, yêu cầu về tính minh bạch và giám sát của con người, với mức phạt vi phạm lên tới 6% doanh thu toàn cầu (Almada & Petit, 2025). Hoa Kỳ ban hành Lệnh Hành pháp về AI (2023) nhằm tạo sự cân bằng giữa đổi mới và an toàn, thiết lập Viện An toàn AI, cấp 500 triệu USD cho nghiên cứu AI có trách nhiệm và khuyến khích các tập đoàn công nghệ lớn cam kết tự nguyện (Boone, 2024). Trung Quốc (2023) tăng cường quản lý các thuật toán và dữ liệu, yêu cầu được phê duyệt AI trong nghiên cứu và kết hợp giá trị xã hội chủ nghĩa với hệ thống tín nhiệm xã hội (Filimowicz, 2023). Các mô hình này thể hiện sự khác nhau trong ưu tiên và phương pháp quản lý AI toàn cầu.

ASEAN đang triển khai nhiều chính sách nhằm thúc đẩy chuyển đổi số khu vực, bao gồm việc xây dựng hướng dẫn đạo đức về trí tuệ nhân tạo (AI ethics guidelines) để các quốc gia thành viên phát triển và ứng dụng AI một cách có trách nhiệm và minh bạch. Đồng thời, ASEAN triển khai các chương trình tăng cường năng lực (capacity building programs) nhằm đào tạo và chuẩn bị nguồn nhân lực cho nền kinh tế số. Ngoài ra, tổ chức này cũng thúc đẩy hợp tác trong quản trị dữ liệu xuyên biên giới (cross-border data governance), nhằm bảo đảm việc lưu chuyển dữ liệu an toàn, tôn trọng quyền riêng tư và hỗ trợ thương mại điện tử. Cuối cùng, ASEAN khuyến khích thử nghiệm công nghệ mới thông qua hộp cát đổi mới sáng tạo (innovation sandboxes) - môi trường thử nghiệm có kiểm soát dành cho các sáng kiến và mô hình kinh doanh mới, đảm bảo tính sáng tạo nhưng vẫn tuân thủ quy định pháp lý (Md Dahlan, 2025).

Việt Nam đang ở giai đoạn đầu trong việc thiết lập khung pháp lý cho AI, với nhiều thiếu sót rõ ràng so với thế giới. EU, Mỹ và Trung Quốc đã xác lập các mô hình quản trị AI với tốc độ và quy mô khác nhau, Việt Nam cần tận dụng lợi thế đi sau để học hỏi, đồng thời đẩy mạnh hành động quyết liệt và đồng bộ nhằm tránh nguy cơ tụt hậu trong cuộc đua toàn cầu về AI.

4.6. Các phát hiện chính và ý nghĩa

Kết quả nghiên cứu chỉ ra một bức tranh đa dạng về việc ứng dụng trí tuệ nhân tạo (AI) trong nghiên cứu khoa học, thể hiện sự thay đổi đáng kể cả về công cụ lẫn cách thức tạo ra tri thức. Theo Giabbanelli và MacEwan (2024), "Chúng ta không chỉ chứng kiến sự thay đổi công cụ mà còn là một sự chuyển biến paradigm trong phương thức sản xuất tri thức". Ba phát hiện chính của nghiên cứu làm nổi bật những xu hướng và thách thức quan trọng. Đầu tiên, tốc độ phát triển AI nhanh hơn khả năng kiểm soát: có nhiều nhà nghiên cứu áp dụng AI, nhưng số lượng tổ chức có quy định cụ thể còn ít ỏi, dẫn đến một lỗ hổng lớn trong quản lý và nguy cơ xuất hiện một "Wild West" trong khoa học, nơi mọi thứ đều có thể xảy ra cho đến khi hậu quả xấu xuất hiện. Thứ hai, lợi ích của AI rất rõ ràng—AI có khả năng nâng cao năng suất lên đến 47%—nhưng nguy cơ vẫn chưa được quản lý một cách toàn diện, biểu hiện qua tỷ lệ ảo tưởng từ 15% đến 40%. Han và đồng nghiệp (2024) nhấn mạnh: "Nếu không cẩn thận, chúng ta có thể đánh đổi chất lượng để lấy số lượng". Thứ ba, sự bất bình đẳng trong khoa học toàn cầu có nguy cơ tăng lên khi tỷ lệ ứng dụng AI ở Bắc Mỹ đạt 81%, trong khi Châu Phi chỉ có 43%, cho thấy khoảng cách số đang dần biến thành khoảng cách AI, làm sâu sắc thêm sự chênh lệch phát triển khoa học giữa các khu vực.

4.7. Khuyến nghị cho Việt Nam

4.7.1. Khuyến nghị về chính sách

Việt Nam cần thiết thiết lập Chiến lược AI cho Nghiên cứu Khoa học giai đoạn 2025-2030 ở cấp quốc gia, với mục tiêu đến năm 2030 có 80% nhà nghiên cứu áp dụng AI một cách có trách nhiệm, phân bổ ngân sách tối thiểu 0,1% GDP cho cơ sở hạ tầng AI và thiết lập hệ thống giám sát KPI cụ thể. Cùng với đó, việc thiết lập Ủy ban Đạo đức AI Quốc gia với sự tham gia đa lĩnh vực và quyền đưa ra khuyến nghị có tính bắt buộc sẽ tạo điều kiện cho quản lý rõ ràng. Cần thiết có một khung pháp lý AI trong giáo dục, quy định rõ ràng về tính minh bạch, trách nhiệm, quyền sở hữu trí tuệ và xử lý các vi phạm. Chương trình AI dành cho tất cả nhà nghiên cứu cần cung cấp công cụ cơ bản miễn phí, giao diện tiếng Việt và hợp tác với các công ty công nghệ. Cuối cùng, Quỹ Nghiên cứu AI Có Trách nhiệm cần được thành lập để tài trợ cho các dự án AI có trách nhiệm và hỗ trợ các tổ chức gặp khó khăn.

Tại cấp tổ chức, các trường và viện cần thiết lập chính sách AI nội bộ bắt buộc trong 6 tháng, thành lập ủy ban đạo đức AI để đánh giá nghiên cứu, đầu tư vào cơ sở hạ tầng (phòng thí nghiệm AI, cụm GPU, phần mềm bản quyền) và đội ngũ chuyên môn. Tại cấp cá nhân, nhà nghiên cứu cần thực hiện danh sách kiểm tra tự đánh giá, cam kết với việc học tập liên tục (20 giờ/năm), tham gia vào mạng lưới hỗ trợ đồng nghiệp và ký kết vào bản cam kết đạo đức.

4.7.2. Khuyến nghị về đào tạo và nâng cao năng lực

Nghiên cứu đề xuất một chương trình đào tạo đa cấp bao gồm: kiến thức về

AI (40 giờ cơ bản, 60 giờ nâng cao), kết hợp AI trong chương trình thạc sĩ và tiến sĩ, cùng với hệ thống chứng chỉ (cơ bản, nâng cao, huấn luyện viên). Tài liệu cần thiết bao gồm Sổ tay AI dành cho nhà nghiên cứu Việt Nam, Trung tâm Tài nguyên Trực tuyến và dự án dịch thuật các hướng dẫn, tài liệu quốc tế. Xây dựng cộng đồng thực hành thông qua mạng lưới VAIRN, các nhóm chuyên đề và chương trình hợp tác sẽ khuyến khích chia sẻ và đổi mới.

4.7.3. Khuyến nghị về cơ sở hạ tầng và công nghệ

Việt Nam cần xây dựng Vietnam AI Research Cloud (VARC) ở cấp quốc gia với khả năng tính toán mạnh mẽ, bộ công cụ AI bản quyền và mã nguồn mở, cùng với cơ sở dữ liệu nghiên cứu được tiêu chuẩn hóa. Tối thiểu mỗi tổ chức cần có AI lab, GPU cluster, phần mềm chính hãng và đội ngũ kỹ thuật. Việt Nam nên tập trung vào việc phát triển các công cụ ngôn ngữ tiếng Việt (trợ lý viết, phân tích tài liệu, trình quản lý trích dẫn, phát hiện đạo văn) cùng với các công cụ chuyên ngành (y tế, nông nghiệp, xã hội, giáo dục).

V. Kết luận

Nghiên cứu đã xem xét một cách toàn diện việc áp dụng AI một cách có trách nhiệm trong lĩnh vực khoa học dựa trên việc tổng hợp 287 tài liệu chất lượng. Kết quả cho thấy: (1) AI đã trở thành công cụ quan trọng với 73% nhà nghiên cứu toàn cầu sử dụng; (2) AI mang lại lợi ích rõ ràng như tăng năng suất 47%, cải thiện chất lượng 35%, và mở rộng cơ hội cho các nhà nghiên cứu ở nước đang phát triển, nhưng cũng đặt ra thách thức về đạo đức, độ tin cậy và sở hữu trí tuệ; (3) các mô hình quốc tế nhấn mạnh bốn nguyên tắc chính: minh bạch, kiểm soát con người, công bằng và trách nhiệm, trong đó EU dẫn đầu về quy định; (4) Việt Nam đang ở giai đoạn đầu nhưng có tiềm năng lớn nhờ vào dân số trẻ, chính sách hỗ trợ và kinh nghiệm số hóa.

Nghiên cứu mang lại 3 đóng góp to lớn sau: (1) phát triển khung lý thuyết tích hợp về AI có trách nhiệm trong nghiên cứu khoa học; (2) cung cấp bộ khuyến nghị toàn diện, thực tiễn, phù hợp bối cảnh Việt Nam; (3) áp dụng phương pháp tổng hợp hệ thống kết hợp phân tích nội dung làm tiền lệ cho các nghiên cứu sau.

AI có trách nhiệm là vấn đề công nghệ, văn hóa và chính sách. Việt Nam cần chiến lược đồng bộ và hành động quyết liệt để tận dụng AI phục vụ phát triển khoa học và bền vững.

Tài liệu tham khảo:

[1]. Almada, M., & Petit, N. (2025). The EU AI Act: Between the rock of product safety and the hard place of fundamental rights. *Common Market Law Review*, *62*(1).

[2]. Avanzo, M., Stancanello, J., Pirrone, G., Drigo, A., & Retico, A. (2024). The evolution of artificial intelligence in medical imaging: from computer science to machine and deep learning. *Cancers*, *16*(21), 3702.

[3]. Austin, E. K., Raile, E. D., Wallner, M. P., Peterson, J., Lewandowski, B., Sellegren, B., ... & Jorgensen, C. (2024). Broadening the concept of value in science and technology innovation policy: Reconsidering cooperative Research and Development Agreements as an expression of Public Value Governance. *Public Administration Quarterly*, 07349149241256037.

[4]. Bisoyi, A. (2022). Ownership, liability, patentability, and creativity issues in artificial intelligence. *Information Security Journal: A Global Perspective*, *31*(4), 377-386.

[5]. Boone, T. S. (2024). An Examination Of Certain Key Features Of The New White House Executive Order On Artificial Intelligence.

[6]. Bozkurt, A. (2024). GenAI et al. Cocreation, authorship, ownership, academic ethics and integrity in a time of generative AI. *Open Praxis*, *16*(1), 1-10.

[7]. Chelli, M., Descamps, J., Lavoué, V., Trojani, C., Azar, M., Deckert, M., ... & Ruetsch-Chelli, C. (2024). Hallucination rates and reference accuracy of ChatGPT and Bard for systematic reviews: comparative analysis. *Journal of medical Internet research*, *26*, e53164.

[8]. Daugherty, P. R., & Wilson, H. J. (2024). *Human+ Machine, Updated and Expanded: Reimagining Work in the Age of AI*. Harvard Business Press.

[9]. Filimowicz, M. (Ed.). (2023). *China's Digital Civilization: Algorithms and Society*. Taylor & Francis.

[10]. Giabbanelli, P. J., & MacEwan, G. (2024). Leveraging artificial intelligence and participatory modeling to support paradigm shifts in public health: an application to obesity and evidence-based policymaking. *Information*, *15*(2), 115.

[11]. Goh, W. W., Chia, K. Y., Cheung, M. F., Kee, K. M., Lwin, M. O., Schulz, P. J., ... & Sung, J. J. (2024). Risk perception, acceptance, and trust of using AI in gastroenterology practice in the Asia-Pacific region: Web-based survey study. *JMIR AI*, *3*(1), e50525.

[12]. Han, W., Xu, J., Cheng, Q., Zhong, Y., & Qin, L. (2024). Robo-Advisors: Revolutionizing Wealth Management Through the Integration of Big Data and Artificial Intelligence in Algorithmic Trading Strategies. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *3*(3), 33-45.

[13]. Jia, F., Sun, D., & Looi, C. K. (2024). Artificial intelligence in science education (2013-2023): Research trends in ten years. *Journal of Science Education and Technology*, *33*(1), 94-117.

[14]. Krippendorff, K. (2023). A critical cybernetics. *Constructivist Foundations*, *19*(1), 82-93.

[15]. Kumer, S. A., Kanakaraja, P., Nadipalli, L. S., Ramesh, N. V. K., & Kotamraju, S. K. (2021, May). The Categorization of Artificial Intelligence (AI) Based. In *Proceedings of International Conference on Communication and Artificial Intelligence: ICCAI 2020* (Vol. 192, p. 411). Springer Nature.

[16]. Lincoln Y. S., & Guba E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.

[17]. Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., ... & Ge, B. (2023). Summary of chatgpt-related research and perspective towards the future of large language models. *Meta-radiology*, *1*(2), 100017.

[18]. Md Dahlan, N. H. (2025). Can ASEAN Build A Sustainable Data Center Future?.

[19]. MarketsandMarkets. (2020). AI Governance Market by Component, Deployment Mode, Organization Size, Vertical and Region - Global Forecast to 2025. Retrieved from https://www.marketsandmarkets.com/industry-analysis/artificial-intelligence-regulatory-compliancemarket

[20]. Markowitz, D. M., Boyd, R. L., & Blackburn, K. (2024). From silicon to solutions: AI's impending impact on research and discovery. *Frontiers in Social Psychology*, *2*, 1392128.

[21]. Martius, F., Jansen, L., Struck, L., Arumugam, A., Geierhaas, L., Ortloff, A. M., ... & Tiefenau, C. (2025, April). Out of Sight, Out of Mind? Exploring Data Protection Practices for Personal Data in Usable Security & Privacy Studies. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1-16).

[22]. Maslej, N., Fattorini, L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., ... & Perrault, R. (2023). Artificial intelligence index report 2023. *arXiv preprint arXiv:2310.03715*.

[23]. Nasim, S. F., Ali, M. R., & Kulsoom, U. (2022). Artificial intelligence incidents & ethics a narrative review. *International Journal of Technology Innovation and Management (IJTIM)*, *2*(2), 52-64.

[24]. Nolan, A. (2024). Artificial intelligence in science: challenges, opportunities and the future of research.

[25]. Nucleoo.com (n.d). The impact of AI on job quality and business efficiency. https://www.nucleoo.com/en/blog/the-impact-of-ai-on-job-quality-and-business-efficiency/

[26]. OECD. (2019). OECD AI Principles. Retrieved from https://www.oecd.org/goingdigital/ai/principles/

[27]. Pho, H. (2024). Investigating Acceptance to Conversational Artificial Intelligence: A Case Study in Vietnam. In *Navigating the Circular Age of a Sustainable Digital Revolution* (pp. 357-380). IGI Global.

[28]. Pulari, S. R., & Jacob, S. G. (2025). Research Insights on the Ethical Aspects of AI-Based Smart Learning Environments: Review on the Confluence of Academic Enterprises and AI. *Procedia Computer Science*, *256*, 284-291.

[29]. Rai, D. H. (2024). Artificial Intelligence Through Time: A Comprehensive Historical Review.

[30]. Sharma, A. (2024). *Advancing AI Understanding in Language & Vision*. University of California, Santa Barbara.

[31]. Somers, M. (2023). How generative AI can boost highly skilled workers' productivity. https://mitsloan.mit.edu/ideas-made-to-matter/how-generative-ai-can-boost-highly-skilled-workers-productivity

[32]. Thaariq, Z. Z. A., Surahman, E., Jaradat, M. S. K., & Mellyana, I. M. (2025). Trends in Using Artificial Intelligence in Social and Natural Science Research. *TAM Akademi Dergisi*, *4*(1), 101-132.

[33]. Thayyib, P. V., Mamilla, R., Khan, M., Fatima, H., Asim, M., Anwar, I., ... & Khan, M. A. (2023). State-of-the-art of artificial intelligence and big data analytics reviews in five different domains: A bibliometric summary. *Sustainability*, *15*(5), 4026.

[34]. Thelwall, M., Jiang, X., & Bath, P. A. (2025). Estimating the quality of published medical research with ChatGPT. *Information Processing & Management*, *62*(4), 104123.

[35]. Trương, H. (2023). ChatGPT và giáo dục: Tổng quan tình hình nghiên cứu trên thế giới và Việt Nam.

[36]. United Nations Educational, Scientific and Cultural Organization, & Organization, C. (2021). *Recommendation on the Ethics of Artificial Intelligence*. United Nations Educational. https://unesdoc.unesco.org/ark:/48223/pf0000379920.page=14

[37]. UNESCO International Research Centre on Artificial Intelligence. (2024). *Challenging systematic prejudices: An investigation into bias against women and girls in large language models*. https://unesdoc.unesco.org/ark:/48223/pf0000388971

[38]. Van der Elst, K. (2025). Foresight on AI: Policy considerations.

[39]. van Kolfschooten, H., & van Oirschot, J. (2024). The EU artificial intelligence act (2024): implications for healthcare. *Health Policy*, *149*, 105152.

[40]. Valavanidis, A. (2023). Artificial intelligence (ai) applications. *Department of Chemistry, National and Kapodistrian University of Athens, University Campus Zografou*, *15784*.

[41]. Voss, W. G. (2016). European union data privacy law reform: General data protection regulation, privacy shield, and the right to delisting. *The Business*

Lawyer, *72*(1), 221-234.

[42]. Wang, X., Li, B., Song, Y., Xu, F. F., Tang, X., Zhuge, M., ... & Neubig, G. (2024). Openhands: An open platform for ai software developers as generalist agents. *arXiv preprint arXiv:2407.16741*.

[43]. Wong, L. H., & Looi, C. K. (2024). Advancing the generative AI in education research agenda: Insights from the Asia-Pacific region. *Asia Pacific Journal of Education*, *44*(1), 1-7.

[44]. Yu, J., & Cooper, H. (1983). A quantitative review of research design effects on response rates to questionnaires. *Journal of Marketing research*, *20*(1), 36-44.

[45]. Zhang, H., Lee, I., & Moore, K. (2023). Preparing teachers to teach artificial intelligence in classrooms: An exploratory study. In *Proceedings of the 17th International Conference of the Learning Sciences-ICLS 2023, pp. 974-977*. International Society of the Learning Sciences.

[46]. Zhang, H., Lee, I., Ali, S., DiPaola, D., Cheng, Y., & Breazeal, C. (2023). Integrating ethics and career futures with technical learning to promote AI literacy for middle school students: An exploratory study. *International Journal of Artificial Intelligence in Education*, *33*(2), 290-324.

RESPONSIBLE USE OF ARTIFICIAL INTELLIGENCE IN SCIENTIFIC RESEARCH AND PUBLICATION: LEGAL PERSPECTIVES FROM CHATGPT TO NEXT-GENERATION AI TOOLS

Nguyen Thanh To[4], Dam Thi Diem Hanh[5]*, Nguyen Vinh Huy[5]

Abstract: *AI is revolutionizing scientific research, bringing both opportunities and challenges. It enhances research efficiency, improves the quality of data analysis, and expands access to information. However, issues related to ethics, transparency, intellectual property rights, and technological risks must be strictly managed. This study examines the responsible use of artificial intelligence (AI) in scientific research and publication, with a focus on developments from ChatGPT to the next generation of AI tools. By analyzing models from the EU, the US, and several ASEAN countries, and assessing the current situation in Vietnam, the article draws lessons for Vietnam in the context of integration and digital transformation.*

Keywords: *responsible AI, research integrity, academic publishing, governance frameworks, large language models*

---

[4] Lecturer at Hoa Sen University

[5] Hanoi Open University

[3] Lecturer at Van Hien University